Engineering management 12 (1) (2026) 12-22

Engineering management

MACHINE LEARNING-BASED PREDICTION OF CAMPUS PLACEMENTS USING ACADEMIC PERFORMANCE AND EMPLOYABILITY SKILLS*

Mukil Kumar Yogesh

National Institute of Technology, Puducherry, India

Abstract

Securing employment after graduation is a primary objective for students pursuing higher education. The increasing competition in the job market has encouraged educational institutions to strengthen their placement strategies by providing students with academic support and employability-focused training. This study aims to predict students' placement outcomes by considering both academic performance and selected psychological characteristics. The factors examined include academic achievement, self-esteem, confidence, emotional intelligence, and anxiety levels, as these attributes play a significant role in employability and can be enhanced through appropriate training interventions.

Keywords: *academic performance, campus placement, employability skills, graduates, predictive analytics*

1. INTRODUCTION

Placement is defined as getting a job through the educational institution during or at the end of the course duration. It is an important milestone in a student's academic journey. Students enroll for higher studies in colleges majorly aim to get a job so that they can apply their knowledge and skills and start earning for their livelihood. Educational institutions train and equip their students to get the best possible job with a high salary package in reputed organizations. Development of employability skills is essential for the career growth of the students. Hence, placement training is taken with utmost care and caution by the academic institutions. Placement training develops students with the required skills and capabilities that are sought after by companies.

College campus placements are platforms where industry’s human resource needs and demands meet with the hopeful aspirations of many promising students. In general, it might be observed that even candidates having similar academic credentials and skills do not get placed for the same company offering the same value propositions in terms of both monetary and non-monetary benefits. This implies that in placing the students in reputed organizations besides the academic credentials many other factors play important role. These factors are considered as psychological factors that could be trained and improved through a structured training process.

The reputation of the educational institute, especially the higher education institutes is a deciding factor for the choice of the students to aspire for it. The reputation is built on

* *Paper presented at the XXII Student Symposium on Strategic Management*

many factors like curriculum, teachers' profile and their skills and knowledge, teaching methodology, infrastructure, programs offered, and mainly the placement records. Students look for these factors before enrolling and they give importance to the placement records of the institute as getting an employment is a major priority for the students. Educational institutes also place a major concern in enhancing the employability opportunities of the students. In doing so, they try to enhance the skills and capabilities of the students which could attract better employment opportunities. With numerous skills and capabilities playing a role in determining the employability, the placement training is unclear on what skills do play a role in determining the employability. With the advent of machine learning, the prediction of employability is made easy.

This is the era of artificial intelligence (AI) in general and machine learning (ML) in particular. Hence, the prediction models based on the machine learning algorithms: Naïve Bayes, support vector machine, random forest, decision tree and logistic regression have been used in this study. The prediction models developed in this work help the educational institutions to take appropriate initiatives to increase the percentage of students placed in reputed organizations. Over a period of time the educational institutions have observed that it is not sufficient to train the students in terms of cumulative grade point average (CGPA) or percentage of marks alone. The placement history indicates that some students with very good CGPA are not able to make it during the placement and some students who seem to be mediocre in terms of academic performance perform well during the placement. The proposed work analyses the role of CGPA and the non-academic attributes like confidence, anxiety, self-esteem, calmness, self-control, self-efficacy, happiness and emotional intelligence, to predict the probability with which students will get the placement.

The collected data from the students are applied to the developed prediction models (that are based on machine learning) and the role of each non-academic attribute in placement is analysed. This analysis enables the educational institutions to identify the strength and weakness of students in various non-academic attributes. Besides analysing the role of each non-academic attribute, the suitability of the machine learning algorithms is also analysed. Once an institution identifies the strength and weakness of the students in terms of non-academic attributes with the help of the developed machine learning algorithms it can plan for counselling, soft skill development and other strategies that will equip the students to get placed in the organizations with better packages that match their knowledge, skill set, intelligent quotient (IQ) and academic performance. In other word this study helps the educational institutions to make their students to overcome the handicaps they have in the non-academic attributes.

2. LITERATURE REVIEW

Majority of the available literature has taken academic performance as a success factor for students and have attempted using machine learning algorithms in the prediction process. Studies have used academic achievement either on the program or on a particular subject as the target variable and models were analyzed using ML algorithms (Rastrollo-Guerrero et al., 2020). Zabriskie, Yang, DeVore and Stewart (2019) have used Naïve Bayesian classifier to classify the students in to a dichotomous group of 'risky students' and 'successful student'. In their study the classification is based on the student behavior-based attributes like student interaction in the class, class login, and task accomplishment.

Use of many Machine Learning Algorithm in the prediction of student performance, in terms of academic performance is reported in few studies. Algorithms like Decision Tree, Support Vector Machines, Naïve Bayesian Classifier, Random Forest are used in predicting student performance (Imran et al., 2019; Sekeroglu et al., 2019). Recently, studies have used artificial neural networks (ANN) to predict student performance (Rodríguez-Hernández et al., 2021). Student's performance is assessed using skills acquired, knowledge acquired, change in behavior and expertise in making decisions. Usage of only one or two algorithms (Sandra et al., 2021) in student performance prediction was common earlier and, in the usage, and comparison of multiple Machine Learning algorithms has been noted recently (Batool et al., 2023; Chango et al., 2021; Tomasevic et al., 2020).

Studies attempting student placement has widely used the available data in the educational institute to predict the employment probabilities of students (Joy & Raj, 2019). Surya et al. (2022) have used past placement percentage of a university to predict the next year's placement. Few supervised machine learning classifiers were used to predict the placement of a student in the IT industry based on their academic performance in class Tenth, Twelve, Graduation, and Backlog till date in Graduation have been proposed by Maurya et al. (2021). The classification algorithms like Support Vector Machine, Gaussian Naive Bayes, K-Nearest Neighbor, Random Forest, Decision Tree, Stochastic Gradient Descent, Logistic Regression, and Neural Network are used to develop the classifiers. They have used accuracy scores, heat map and classification reports to test the accuracy of the various tested algorithms. They reported SVM and logistic regression to show high accuracy when compared to other algorithms. Cakit and Dagdeviren (2022) had attempted a similar study but with input parameters of educational institutes' facilities, location and reputation and found that of the various algorithms used extreme gradient boosting (EGB) algorithm resulted in better accuracy. This study is done with Turkish universities.

It has been observed from our survey that performance and other student related achievements are analyzed only in terms of academic attributes. Many of them consider only academic attributes and only very few non- academic attributes or no non-academic attribute at all. Hence, an attempt has been made to estimate the probability of a student getting placed in terms of a set of relevant academic and non-academic attributes. It is proven that non-academic and psychological attributes play a significant role in the success of a students and especially the placement (Norida et al., 2014; Belle et al., 2022; Ristiani & Lusianingrum, 2022; Karvendan & Jayakumar, 2025). Hence, this study has taken psychological factors namely confidence, anxiety, self-esteem, calmness, self-control, self-efficacy, happiness and emotional intelligence as predictors of student placement.

3. DATA AND METHODOLOGY

3.1. Data collection

For the CGPA and non-academic attributes, data was collected from final year students of engineering colleges and universities using a structured questionnaire. All the attributes were measured using standard questionnaire. For example, Self-esteem was measured using 10 item Rosenberg scale (1965), emotional intelligence (EI) was measured by brief emotional intelligence questionnaire (BEIS-10) (Davies et al., 2010). The questionnaire was distributed through Google forms to various educational institutions and gathered the data across educational institutions. After reminders and request for forwards, the data was collected for a period of 3 months. The data set has 830 data points. The data set includes the following attributes: CGPA, confidence, anxiety, self-esteem, calmness, self-control, self-efficacy, happiness, emotional intelligence and placement. Placement is the target variable, taken as dichotomous variable 'placed' and 'not placed'. Once the data is collected, they are cleaned and pre-processed. The data set has 830 data points. As a thumb rule, we have used 70% of the collected data for training and 30% for testing.

3.2. ML based Prediction models

Since the data set has so many attributes apart from CGPA it is very difficult to analyze the probability of a student getting placed by making an exhaustive analysis or a brute force method. Machine learning (ML) algorithms which have been formulated based on Artificial Intelligence (AI) comes to our help to do this challenging task. In this study Naïve Bayes, support vector machine, random forest, and logistic regression machine learning algorithms are used.

3.3. Proposed Models

All machine learning algorithms build a prediction model and a Model is an explicit description of patterns within the data. Models based on machine learning algorithms have three associated parameters namely task, experience and performance measure. All the machine learning algorithms used in this study come under the category of supervised learning since they expect the data points to be labelled. The data set is a labelled data set with dichotomous labelling of ‘placed’ and ‘not placed’.

Naive Bayes is based on Baye’s Theorem, an approach to calculate conditional probability based on prior knowledge, and the naive assumption that each feature is independent to each other. The biggest advantage of Naive Bayes is that, while most machine learning algorithms rely on large amount of training data, it performs relatively well even when the training data size is small.

Support vector machine (SVM) finds the best way to classify the data based on the position in relation to a border between positive class and negative class. This border is known as the hyperplane which maximize the distance between data points from different classes. Similar to decision tree and random forest, support vector machine can be used in both classification and regression, support vector classifier (SVC) is for classification problem.

Random forest is a collection of decision trees. It is a common type of ensemble methods which aggregate results from multiple predictors. Random forest additionally utilizes bagging technique that allows each tree trained on a random sampling of original dataset and takes the majority vote from trees. Compared to decision tree, it has better generalization but less interpretable, because of more layers added to the model.

Logistics regression uses sigmoid function above to return the probability of a label. It is widely used when the classification problem is binary, true or false, win or lose, positive or negative. In our study, it is whether a student has got placed or not. Dependent Variable is binary. The sigmoid function generates a probability output. By comparing the probability with a pre-defined threshold, the object is assigned to a label accordingly.

In this study, these four machine learning based prediction models that predict the probability with which the students will secure a job are proposed and implemented. The prediction models have been implemented in Python programming language. Python is a popular and versatile programming language. The proposed models have been evaluated in terms of a set of prediction evaluation metrics.

3.4. Prediction evaluation metrics

All the four prediction models need to be evaluated to decide the suitability of these models for the chosen problem. Besides deciding their suitability, it is also necessary to suggest which one among the four is more suitable. To evaluate the proposed models a set of prediction evaluation metrics is required. This section explains the various performance evaluation metrics used in this study.

3.4.1. Confusion Matrix

Confusion Matrix is a matrix that describes the complete performance of the model. When performing classification predictions, there are four types of outcomes that could occur.True positives are when you predict an observation belongs to a class and it actually does belong to that class. True negatives are when you predict an observation does not belong to a class and it actually does not belong to that class. False positives occur when you predict an observation belongs to a class when in reality it does not. False negatives occur when you predict an observation does not belong to a class when in fact it does.

3.4.2. Accuracy

Accuracy is the fraction of predictions our model got right. Accuracy is the proportion of number of correct predictions to the total number of predictions. It works well only if there are equal number of samples belonging to each class.

3.4.3. Precision

It is the number of correct positive results divided by the number of positive results predicted by the classifier.

3.4.4. Recall

It is the number of correct positive results divided by the number of *all* relevant samples (all samples that should have been identified as positive).

3.4.5. F score

The range for F score is [0, 1]. It tells how precise the classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances). The greater the F1 Score, the better is the performance of the model.

3.4.6. Receiver Operating Characteristics (ROC)

Receiver operating characteristics is the area under curve, which is one of the most widely used metrics for evaluation. It is used for binary classification problem. ROC of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example. ROC is the area under the curve of plot *False Positive Rate vs True Positive Rate* at different points in [0, 1]**.** ROC has a range of [0, 1]. The greater the value, the better is the performance of the model.

4. RESULTS AND DISCUSSION

The test for normal distribution of the collected dataset is check for its suitability in applying it to the various prediction models. Except the predictor variable, placement, all other attributes are checked for normal distribution of datasets. Shapiro- Wilk test was applied and it was observed that all the attributes have the significance $p>0.05$, indicating that all the attributes are normally distributed and hence, we can apply the chosen ML algorithms.

The prediction models based on Naïve Bayes algorithm, support vector machine, random forest and logistic regression have been implemented using Python programming language. The performance measure of the Naïve Bayes algorithm is given in Figures 1 and 2.

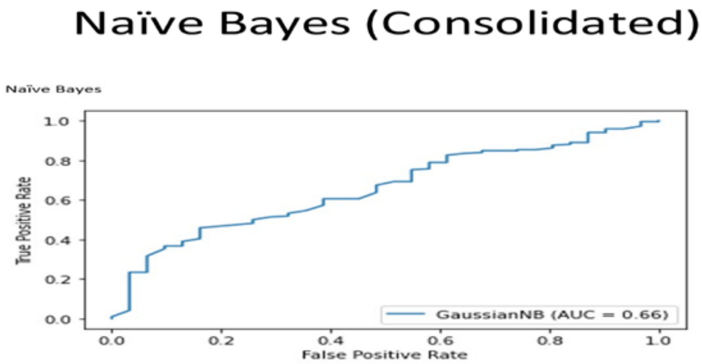

Figure 1. Naïve Bayes - True positive & False positive

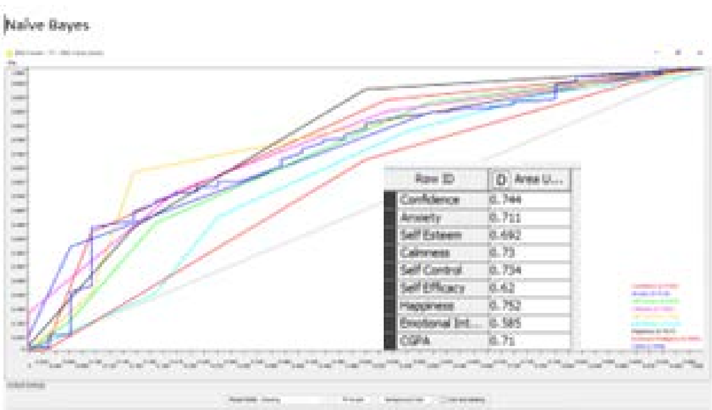

Figure 2. Naïve Bayes - ROC & area under curve

Table 1 presents a comparative overview of the accuracy statistics obtained in the analysis, enabling a clear evaluation of model performance across different approaches.

Table 1. Comparative Results - accuracy statistics

Algorithm	Accuracy	Recall	Precision	F-Score	Specificity	Sensitivity
SVM	0.992	0.991	0.926	0.962	0.991	0.991
Logistic regression	0.855	0.972	0.875	0.921	0.91	0.972
Naïve Bayes	0.996	0.995	0.969	0.998	0.995	0.995
Random Forest	0.859	0.973	0.881	0.924	0.973	0.973

Similar performance measures were taken for all algorithms and the consolidated results are given in table. From Table 1, it is inferred that Naïve Bayes algorithm gives better results when compared to the other algorithms, Random Forest, SVM and Logistics Regression. The consolidated results of the area under the curve of the attributes considered are given in Table 2.

Table 2. Comparative results - Area under curve

Attributes					
Algorithm	Self -control	Self -Efficacy	Happiness	EI	CGPA
SVM	0.59	0.56	0.6	0.626	0.6
Logistic regression	0.718	0.696	0.7	0.644	0.7
Random Forest	0.698	0.618	0.6	0.564	0.6
Naïve Bayes	0.734	0.62	0.8	0.585	0.7

From Table 2 it could be inferred that the area occupied by the various attributes considered is more for naïve Bayes algorithm, when compared to the other ML algorithms. It could also be inferred that the non- academic attributes play a vital role in predicting the placement of the students. Attributes like happiness, self-control, confidence, self-esteem plays a predominant role in predicting the placement of students when compared to CGPA.

From this study it is found that Naïve Bayes is a better algorithm in predicting the student placement when compared to the other algorithms like SVM, Random Forest and Logistic Regression. Naive Bayes algorithm performs well for working with text classification. As the placement is taken as a text “placed” and “not placed”, Naïve Bayes has been effective. Moreover, Naïve Bayes algorithm works well for a small data set with limited number of attributes. The data set used for this study is also small in number with few attributes, it might result in Naïve Bayes being a better prediction algorithm

The present study concludes that psychological factors, self-control, self-efficacy, psychological well-being, and emotional intelligence exert a stronger influence on employability than academic performance measured through CGPA. This finding highlights the importance of non-cognitive competencies as critical determinants of getting a job, a measure of career success. While CGPA reflects academic knowledge and cognitive ability, it does not fully capture the interpersonal, emotional, and behavioral competencies required for successful job attainment and performance.

Previous research has also established that self-efficacy mediates the relationship between emotional intelligence and employability, suggesting that individuals who are emotionally intelligent are more confident in their abilities, which in turn improves their employment prospects (Dacre et al., 2013). Individuals with higher levels of psychological well-being, which is measured through the attribute happiness, are better equipped to manage stress, maintain focus, and exhibit a positive attitude during the recruitment and selection process. Higher education institutions traditionally emphasize academic achievement, often measured through CGPA. However, the labor market demands a broader set of competencies, including emotional intelligence, adaptability, and resilience. The present study supports the argument that educational institutions should integrate psychological skill development into their curricula to better prepare students for employment. This finding will help the trainers, teachers and the students in developing the psychological abilities of the students to fetch employment.

5. CONCLUSION

Education is considered a vital need for motivating self-assurance as well as providing the things that are needed to partake in today's World. Technologies like Artificial Intelligence had a surprising evolution in many fields especially in educational teaching and learning processes. Prediction of student’s performance in terms of fetching a job can be used as a basis for early intervention on the potential failure of students to achieve the goal and to make changes in the strategies of the institutions to facilitate the student achievement.

The major conclusion of this paper is that the psychological variables play a vital role in predicting placement when compared to CGPA. This finding helps educational institutions to realize the need for imparting soft skills to the students apart from the curriculum. From the output of various prediction models, it is found that Naïve Bayes is a better algorithm than other algorithms. It is observed that Naive Bayes algorithm performs well for working with text classification and for the data sets with limited number of attributes.

PREDVIĐANJE ZAPOŠLJAVANJA STUDENATA PRIMENOM MAŠINSKOG UČENJA NA OSNOVU AKADEMSKOG USPEHA I VEŠTINA ZAPOŠLJIVOSTI

Mukil Kumar Yogesh
Nacionalni institut za tehnologiju, Pudučeri, Indija

Izvod

Obezbeđivanje zaposlenja nakon diplomiranja predstavlja jedan od primarnih ciljeva studenata koji pohađaju visoko obrazovanje. Sve veća konkurencija na tržištu rada podstakla je obrazovne institucije da unaprede svoje strategije zapošljavanja studenata pružanjem akademske podrške i obuka usmerenih na razvoj veština zapošljivosti. Cilj ove studije je predviđanje ishoda zapošljavanja studenata uzimajući u obzir kako akademski uspeh, tako i odabrane psihološke karakteristike. Analizirani faktori obuhvataju akademska postignuća, samopoštovanje, samopouzdanje, emocionalnu inteligenciju i nivo anksioznosti, budući da ove karakteristike imaju značajnu ulogu u zapošljivosti i mogu se unaprediti odgovarajućim programima obuke i razvoja.

Ključne reči: *akademski uspeh, zapošljavanje studenata, veštine zapošljivosti, diplomci, prediktivna analitika*

LITERATURA / REFERENCES

Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies, 28(1), 905-971.

Belle, M. A., Antwi, C. O., Ntim, S. Y., Affum-Osei, E., & Ren, J. (2022). Am I gonna get a job? Graduating students' psychological capital, coping styles, and employment anxiety. Journal of Career Development, 49(5), 1122-1136.

Çakıt, E., & Dağdeviren, M. (2022). Predicting the percentage of student placement: A comparative study of machine learning algorithms. Education and Information Technologies, 27(1), 997-1022.

Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. Computers & Electrical Engineering, 89, 106908.

Dacre Pool, L., & Qualter, P. (2013). Emotional self-efficacy, graduate employability, and career satisfaction: Testing the associations. Australian Journal of Psychology, 65(4), 214-223.

Davies, K. A., Lane, A. M., Devonport, T. J., & Scott, J. A. (2010). Validity and reliability of a brief emotional intelligence scale (BEIS-10). Journal of Individual Differences, 31(4).

Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student academic performance prediction using supervised learning techniques. International Journal of Emerging Technologies in Learning, 14(14), 92-104.

Joy, L. C., & Raj, A. (2019, March). A review on student placement chance prediction. In 2019 5th International conference on advanced computing & communication systems (ICACCS), 542-545. IEEE.

Karvendhan, A., & Jayakumar, K. N. (2025). Can Emotional Intelligence, Self-Regulation and Cognitive Styles of College Students Predict Job Placement?. Psychological Thought, 18(1), 173-193.

Maurya, L. S., Hussain, M. S., & Singh, S. (2021). Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. Applied Artificial Intelligence, 35(6), 403-420.

Norida, A., Tajudin, N. M., Kalthom, H., Jano, Z., Sharif, M., & Shahrulanuar, M. (2014, July). Model of self-esteem, job-search intensity and career decision-making self-efficacy for undergraduate students. In Proceedings of the International Conference on Science, Technology and Social Sciences (ICSTSS) 2012, 257-265. Singapore: Springer Singapore.

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. Applied sciences, 10(3), 1042.

Ristiani, R., & Lusianingrum, F. P. W. (2022). The effect of self-confidence on job readiness. Journal of Applied Business, Taxation and Economics Research, 1(5), 450-461.

Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. Computers and Education: Artificial Intelligence, 2, 100018 (1-14).

Rosenberg, M. (1965). Rosenberg self-esteem scale. Journal of Religion and Health.

Sandra, L., & Lumbangaol, F. (2021). Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review. TEM Journal, 10(4), 1919-1927.

Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In Proceedings of the 2019 8th international conference on educational and information technology, 7-11.

Surya, M. S., Kumar, M. S., & Gandhimathi, D. (2022, April). Student placement prediction using supervised machine learning. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 1352-1355. IEEE.

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & education, 143, 103676.

Zabriskie, C., Yang, J., DeVore, S., & Stewart, J. (2019). Using machine learning to predict physics course outcomes. Physical Review Physics Education Research, 15(2), 020120(1-18).